



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing

**Citation for published version:**

Palaiokostas, C, Kocour, M, Prchal, M & Houston, R 2018, 'Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing', *Frontiers in genetics*, vol. 9, 82. <https://doi.org/10.3389/fgene.2018.00082>

**Digital Object Identifier (DOI):**

[10.3389/fgene.2018.00082](https://doi.org/10.3389/fgene.2018.00082)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Frontiers in genetics

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.





# Accuracy of Genomic Evaluations of Juvenile Growth Rate in Common Carp (*Cyprinus carpio*) Using Genotyping by Sequencing

Christos Palaikostas<sup>1</sup>, Martin Kocour<sup>2</sup>, Martin Prchal<sup>2</sup> and Ross D. Houston<sup>1\*</sup>

<sup>1</sup> The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom,

<sup>2</sup> Faculty of Fisheries and Protection of Waters, South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, University of South Bohemia in České Budějovice, Vodňany, Czechia

## OPEN ACCESS

### Edited by:

Riccardo Negrini,  
Università Cattolica del Sacro Cuore,  
Italy

### Reviewed by:

Nicholas Andrew Robinson,  
Nofima, Norway  
Mario Calus,  
Wageningen University & Research,  
Netherlands

### \*Correspondence:

Ross D. Houston  
ross.houston@roslin.ed.ac.uk

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 13 January 2018

**Accepted:** 26 February 2018

**Published:** 13 March 2018

### Citation:

Palaikostas C, Kocour M, Prchal M  
and Houston RD (2018) Accuracy  
of Genomic Evaluations of Juvenile  
Growth Rate in Common Carp  
(*Cyprinus carpio*) Using Genotyping  
by Sequencing. *Front. Genet.* 9:82.  
doi: 10.3389/fgene.2018.00082

Cyprinids are the most important group of farmed fish globally in terms of production volume, with common carp (*Cyprinus carpio*) being one of the most valuable species of the group. The use of modern selective breeding methods in carp is at a formative stage, implying a large scope for genetic improvement of key production traits. In the current study, a population of 1,425 carp juveniles, originating from a partial factorial cross between 40 sires and 20 dams, was used for investigating the potential of genomic selection (GS) for juvenile growth, an exemplar polygenic production trait. RAD sequencing was used to identify and genotype SNP markers for subsequent parentage assignment, construction of a medium density genetic map (12,311 SNPs), genome-wide association study (GWAS), and testing of GS. A moderate heritability was estimated for body length of carp at 120 days (as a proxy of juvenile growth) of 0.33 (s.e. 0.05). No genome-wide significant QTL was identified using a single marker GWAS approach. Genomic prediction of breeding values outperformed pedigree-based prediction, resulting in 18% improvement in prediction accuracy. The impact of reduced SNP densities on prediction accuracy was tested by varying minor allele frequency (MAF) thresholds, with no drop in prediction accuracy until the MAF threshold is set <0.3 (2,744 SNPs). These results point to the potential for GS to improve economically important traits in common carp breeding programs.

**Keywords:** aquaculture breeding, carps, high-throughput sequencing, RAD-seq, genomic prediction

## INTRODUCTION

Carps have the highest global production volume of all aquaculture fish (FAO, 2015) and are farmed in a wide variety of environments and production systems (Balon, 1995). In common with most aquaculture species, only a minority of farmed common carp are derived from family-based selective breeding programs, and crossbreeding of partially inbred strains is commonly applied to benefit from heterosis (Hulata, 1995; Vandeputte, 2003; Janssen et al., 2017). Family-based programs have the advantage of enabling cumulative increases in economic traits, and maintaining a high degree of control of the level of inbreeding of stocks. Empirical data relating to selective breeding of family-based programs in several fish species show an increase in genetic gain of up to 15% per generation (Gjedrem, 2000). While initial studies suggested that within breed selection

is ineffective in common carp (Moav and Wohlfarth, 1976), recent studies focusing on growth and survival traits highlighted the potential of applying selective breeding to enhance production (Kocour et al., 2007; Vandeputte et al., 2008; Nielsen et al., 2010; Ninh et al., 2013; Dong et al., 2015).

Traditional pedigree-based selective breeding incorporating best linear unbiased predictor (BLUP) methodology (Henderson, 1975) has greatly benefited both animal and plant agriculture. Nevertheless, the utilization of just the between-family component of genetic variation imposes limitations to selection accuracy and therefore genetic gain (Meuwissen et al., 2013). Selective breeding can be significantly enhanced by the application of genomic tools, via improvement of selection accuracy and potentially also by identification of causative factors impacting on key production traits (Meuwissen et al., 2001, 2013; Hickey et al., 2017). Genomic selection (GS) involves the simultaneous use of genome-wide genetic markers to estimate breeding values for selection candidates utilizing both within and across family variation (Meuwissen et al., 2001). By using all markers in the calculation of breeding values, GS overcomes the limitations of marker-assisted selection for such traits, where typically only a small percentage of genetic variation can be utilized for polygenic traits (Daetwyler et al., 2013). In aquaculture species, GS has been enabled by the increased technical feasibility and reduced cost of generation of genome-wide marker data in non-model organisms via SNP arrays or genotyping by sequencing (Davey et al., 2011; Robledo et al., 2017).

The effectiveness of GS at deriving more accurate breeding values than traditional pedigree-based selection has been demonstrated using simulated and empirical data in both livestock and aquaculture (Goddard and Hayes, 2009; Sonesson and Meuwissen, 2009). Empirical data collected to date suggest that the majority of traits of interest for animal production (e.g., growth and disease resistance) are underpinned by a polygenic genetic architecture (de los Campos et al., 2013). In aquaculture species, where large full-sibling family sizes are typically available, the advantages of genomic prediction of breeding values in aquaculture species are clear from several studies of such polygenic traits (e.g., Odegård et al., 2014; Tsai et al., 2015, 2016; Vallejo et al., 2017), albeit genomic prediction is less effective when only distant relatives are used in deriving the prediction equation (Tsai et al., 2016).

Restriction-site-associated DNA sequencing (RAD-seq) is a reduced representation of high-throughput sequencing technique for the concurrent detection and genotyping of SNP markers (Baird et al., 2008). RAD-seq and similar genotyping by sequencing techniques rely on digestion of the genomic DNA with a restriction enzyme, and subsequent high-depth sequencing of the flanking regions. These techniques have been widely applied due to their cost-efficiency in a wide range of aquaculture species (Robledo et al., 2017), both in genome-wide association studies (GWAS) (e.g., Campbell et al., 2014; Palti et al., 2015; Barria et al., 2017) and GS studies (e.g., Palaikostas et al., 2016; Vallejo et al., 2016). The main aim of this study

was to investigate the potential of genomic prediction of an exemplar polygenic trait in common carp (juvenile growth) using genome-wide SNP markers generated by RAD-seq. To achieve this, samples of 1,425 carp measured for body weight and length at approximately 4 months of age were used. RAD-seq was used to genotype genome-wide SNP markers, parentage assignment was performed, and heritability (of body weight and length) was estimated. The obtained SNPs were utilized for construction of a medium density linkage map, followed by a GWAS to test the association between individual loci and growth. Finally, GS was tested to evaluate the potential of incorporating genomic data for selective breeding compared to pedigree-based selection using this exemplar polygenic trait.

## MATERIALS AND METHODS

### Sample Collection

A population of Amur Mirror Carp was created at the University of South Bohemia, Czech Republic in May 2014 using artificial insemination (Vandeputte et al., 2004) involving four factorial crosses each comprising 5 dams  $\times$  10 sires (20 dams and 40 sires in total). Incubation of eggs was performed in 9 lt Zugar jars at 20°C. At the swimming stage, randomly sampled progeny of the same total volume from each mating was pooled and reared under semi-intensive pond conditions throughout the growing season (from May to September). In September, a sample of 1,425 fish was fin-clipped, passive integrated transponder (PIT)-tagged, weighed to the nearest 0.01 g, and measured for standard length (SL) (from the tip of the nose to the end of the caudal peduncle) to the nearest millimeter. All working procedures complied with the European Union Directive 2010/63/EU for the protection and welfare of animals used for scientific purposes.

### RAD Library Preparation and Sequencing

Genomic DNA was extracted from fin samples using the REALPure genomic DNA extraction kit (Durviz S.L.) and treated with RNase. Each sample was quantified by spectrophotometry (Nanodrop), and its quality was assessed by agarose gel electrophoresis, before being diluted to a concentration of 20 ng/ $\mu$ L [measured by Qubit Fluorometer (Invitrogen)] in 5 mmol/L Tris, pH 8.5.

The RAD-specific P1 and P2 paired-end adapters and library amplification PCR primer sequences used in this study are detailed in Baxter et al. (2011). Briefly, each sample (0.72  $\mu$ g parental DNA/0.24  $\mu$ g offspring DNA) was digested at 37°C for 60 min with *Sbf*I (recognizing the CCTGCA| GG motif) high fidelity restriction enzyme (New England Biolabs, NEB). The reactions (12  $\mu$ L final volumes) were then heat inactivated at 65°C for 20 min. Individual-specific P1 adapters, each with a unique 5 bp barcode, were ligated to the *Sbf*I-digested DNA at 20°C for 60 min by adding 1.8/0.6  $\mu$ L 100 nmol/L P1 adapter, 0.45/0.15  $\mu$ L 100 mmol/L rATP (Promega), 0.75/0.25  $\mu$ L 10 $\times$  Reaction Buffer 2 (NEB), 0.36/0.12  $\mu$ L T4 ligase (NEB, 2 M U/mL), and reaction volumes made up to 45/15  $\mu$ L

with nuclease-free water for each parental/offspring sample. Following heat inactivation at 65°C for 20 min, the ligation reactions were slowly cooled to room temperature (over 1 h) then combined in appropriate multiplex pools. Shearing and initial size selection (300–600 bp) by agarose gel separation were followed by gel purification, end repair, dA overhang addition, P2 (individual-specific adapters) paired-end adapter ligation, library amplification, as described in the original RAD protocol (Baird et al., 2008; Etter et al., 2011). A total of 150 µL of each amplified library (14–17 PCR cycles, library dependent) was size selected (ca. 400–700 bp) by gel electrophoresis as described in Houston et al. (2012). Following a final gel elution step into 20 µL EB buffer (MinElute Gel Purification Kit, Qiagen), 66 libraries (24 animals each) were sent to BMR Genomics (Italy), for quality control and high-throughput sequencing. Libraries were run in 14 lanes of an Illumina NextSeq 500, using 75 base paired-end reads (v2 chemistry). The sequence reads were deposited at the NCBI Sequence Read Archive (SRA) under the accession PRJNA414021.

## Genotyping RAD Alleles – SNP Identification

Reads missing the restriction site, with ambiguous barcodes and PCR duplicates, were identified and discarded using the Stacks software 1.4 (Catchen et al., 2011). The remaining reads were aligned to the common carp reference genome assembly version GCA\_000951615.2 (Xu et al., 2014) using bowtie2 (Langmead and Salzberg, 2012). The aligned reads were sorted into loci and genotypes using the Stacks software 1.4 (Catchen et al., 2011). A minimum stack depth of at least 10 or 5 was required for the parental and offspring samples, respectively. Only loci containing one or two SNPs were considered for downstream analysis. SNPs with minor allele frequency (MAF) <0.01, >25% missing data, and deviating from expected Hardy–Weinberg equilibrium in the parental samples ( $P < 1e-06$ ) were discarded.

## Parentage Assignment

Due to the partial factorial crosses performed in this experiment, the family structure of the offspring was unknown at the time of sampling. Parentage assignment was performed using the RAD-seq-derived SNP data and the R/hsphase (Ferdosi et al., 2014) software allowing for a maximum overall genotyping error of 4%. The pedigree obtained was further validated for possible erroneous assignments using FImpute (Sargolzaei et al., 2014).

## Linkage Map Construction

Linkage map construction was performed using Lep-Map v2 (Rastas et al., 2013). SNPs with MAF <0.05 in individual families and those deviating from expected Mendelian segregation ( $P < 0.001$ ) were excluded. Linkage groups were formed using a minimum LOD threshold value of 18 in the “SeparateChromosomes” module, allowing a maximum distance between consecutive SNPs of 50 cM. Marker order within

each linkage group was performed using the “OrderMarkers” module using the *SexAverage* option. Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function.

## Heritability Estimation

Heritability estimates of weight and length were performed using both the pedigree-based relationship matrix and the genomic relationship matrix. Variance components was estimated using AIREMLF90 (Miszta et al., 2002) with the following animal model:

$$y = Xb + Zu + e, \quad (1)$$

where  $y$  is the vector of recorded phenotypes,  $b$  vector of the fixed effects (the four-level factorial cross),  $X$  the incidence matrix relating phenotypes with the fixed effects,  $Z$  the incidence matrix relating phenotypes with the random animal effects,  $u$  the vector of random animal effects  $\sim N(0, A\sigma_g^2)$  with  $A$  corresponding to the pedigree-based relationship matrix or the genomic relationship matrix  $G$  (VanRaden, 2008),  $\sigma_g^2$  the additive genetic variance,  $e$  the vector of residuals  $\sim N(0, I\sigma_e^2)$  and  $\sigma_e^2$  the residual variance.

Heritability was estimated using the following formula:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}.$$

Bivariate models with the same fixed and random effects as in Equation (1) were used in order to estimate genetic correlations between the phenotypes of weight and length.

## Genome-Wide Association Analysis

To test the association between individual SNPs and growth (only length records were utilized), GWAS was performed using R/gaston (Hervé and Dandine-Roulland, 2018). The mixed model applied had the same format as in Equation (1) with the addition of each tested SNP as a fixed effect. The genome-wide significance threshold was calculated using a Bonferroni correction ( $0.05/N$ ), where  $N$  represents the number of QC-filtered SNPs.

## Genomic Selection

The accuracy of prediction of genomic breeding values (GEBVs) was calculated and benchmarked against the accuracy of prediction for EBVs using traditional pedigree-based best linear unbiased prediction (BLUP) (Henderson, 1975). GEBVs were estimated with GBLUP (Meuwissen et al., 2001) using the BLUPF90 suite (Miszta et al., 2002) updated for genomic analyses (Aguilar et al., 2011). Pedigree-based BLUP was applied to calculate breeding values using the same software. The general form of the fitted models was as in Equation (1).

A fivefold cross-validation was performed in order to test prediction accuracy. The data set was randomly split into sequential non-overlapping training ( $n = 972$  individuals) and

validation sets ( $n = 242$ ). The fivefold cross-validation procedure was repeated 10 times in order to reduce random sampling effects. The accuracy of the estimated breeding values was approximated as:

$$r = (EBV, y)/h,$$

where  $y$  is the vector of recorded phenotypes and  $h$  the square root of the heritability. GEBVs were used to approximate accuracy in the case of GBLUP. Additionally, five different scenarios were used in order to test the effect of reduced genotyping densities on the obtained prediction accuracies. In these scenarios, GBLUP was performed as above using subsets of SNPs selected by progressive increase of a minimum threshold for MAF. The tested scenarios involved MAF thresholds of 0.1 (8,237 SNPs), 0.2 (4,950 SNPs), 0.3 (2,744 SNPs), 0.4 (1,182 SNPs), and 0.45 (530 SNPs). Bias in the form of the regression coefficient of the phenotypic trait on (G)EBV was estimated for both PBLUP and all tested scenarios of GBLUP.

## RESULTS

### Descriptive Statistics

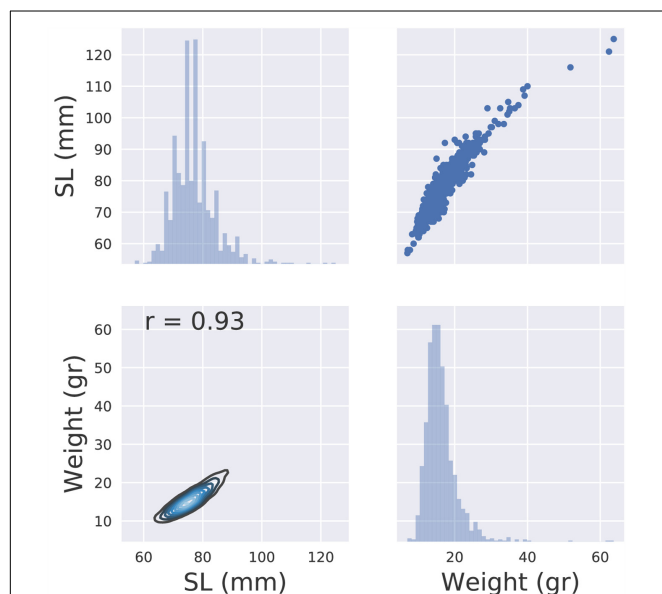
The mean weight of the genotyped carp juveniles after approximately 4 months of growth (**Supplementary Table S1**) was 16.3 g (SD 4.6) and the mean SL was 77 mm (SD 7.1). The Pearson correlation coefficient between length and weight was  $r = 0.93$  (**Figure 1**).

### SNP Identification and Genotyping

Animals with fewer than 25% missing SNP genotypes were retained for downstream analysis (corresponding to 60 parental and 1,400 offspring samples). The total number of raw

**TABLE 1** | Number of QC-filtered SNPs per linkage group.

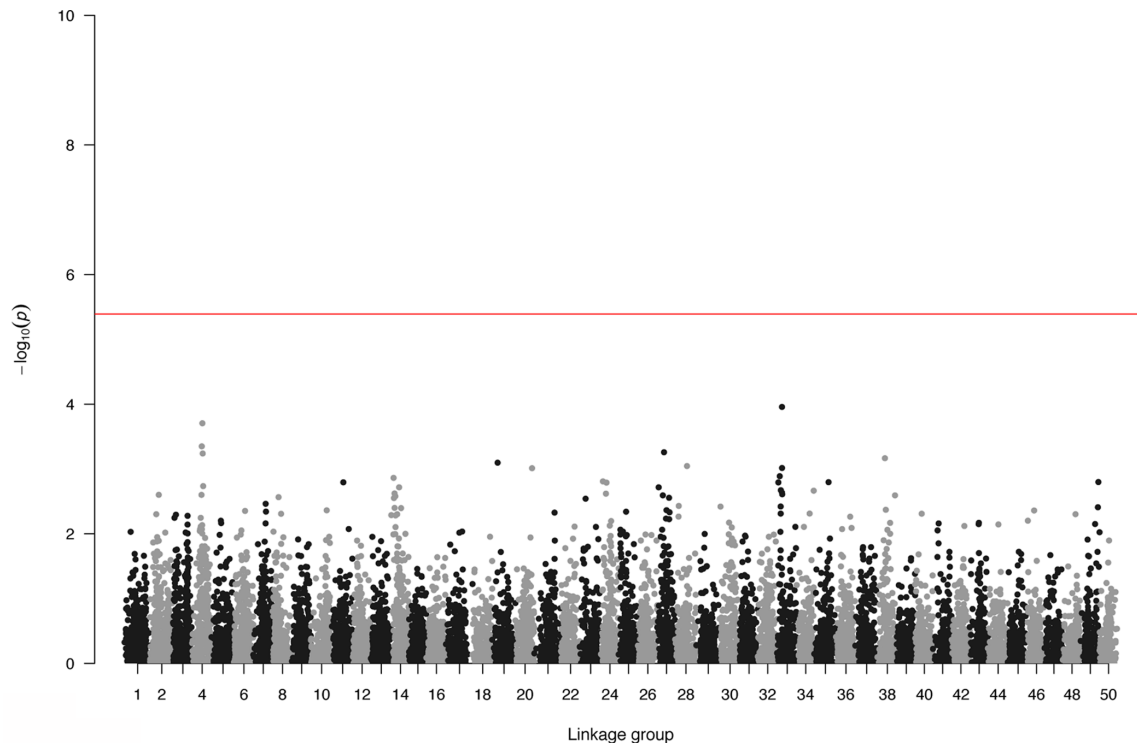
Linkage group	Size (cM)	Number of markers
1	102	350
2	91	350
3	76	325
4	77	322
5	90	319
6	77	315
7	77	311
8	77	292
9	75	290
10	84	287
11	83	283
12	71	277
13	81	272
14	70	263
15	69	260
16	78	259
17	105	257
18	78	256
19	92	249
20	75	249
21	106	246
22	75	245
23	84	241
24	70	241
25	76	240
26	77	239
27	71	238
28	84	237
29	93	237
30	81	236
31	71	229
32	77	229
33	78	227
34	70	227
35	87	226
36	77	225
37	88	222
38	77	221
39	73	215
40	71	213
41	74	212
42	72	212
43	72	208
44	79	203
45	80	199
46	67	193
47	73	175
48	71	169
49	75	163
50	69	157
Total	3,944	12,311



**FIGURE 1** | Descriptive statistics of the phenotypic data (weight, length).

reads passing the QC filters was 6.89 (SD 1.33) M for the parental samples and 3.68 (SD 1.28) M for the offspring. The mean number of RAD loci identified was 57,983 (SD 1,573) and 57,235 (SD 5,224) for parents and offspring, respectively, with mean coverage of 60 (SD 10) X and 29 (SD 8) X, respectively. A total of 22,756 putative SNPs





**FIGURE 2 |** Genome-wide association plot for body length.

were identified, of which 20,039 SNPs passed QC filters and were retained for downstream analysis (**Supplementary Table S2**).

### Parentage Assignment

The carp progeny was assigned to unique parental pairs allowing for a maximum genotypic error rate of 4%. In total 1,214 offspring were uniquely assigned, forming 195 full-sib families (40 sires, 20 dams) ranging from 1 to 21 animals per family with a mean size of 6 (SD 4). The individual dam contribution to the population ranged from 9 to 99 animals with a mean of 61 (SD 23), while the sire contribution ranged from 7 to 53 animals with a mean of 30 (SD 12).

### Linkage Map Construction

The linkage map constructed using the aforementioned families consisted of 12,311 SNPs (**Table 1**) that were grouped into 50 linkage groups (in accordance with the expected karyotype). The length of the consensus linkage map was 3,944 cM (**Supplementary Table S3**). The number of SNPs per chromosome ranged from 157 to 350 (mean = 246; SD = 45), while linkage group length ranged between 67 and 106 cM (mean = 79; SD = 9).

### Heritability Estimation

The estimated heritabilities for weight and length were 0.26 (SE 0.05) and 0.33 (SE 0.05), respectively, and were consistent between the pedigree and genomic models. Genetic

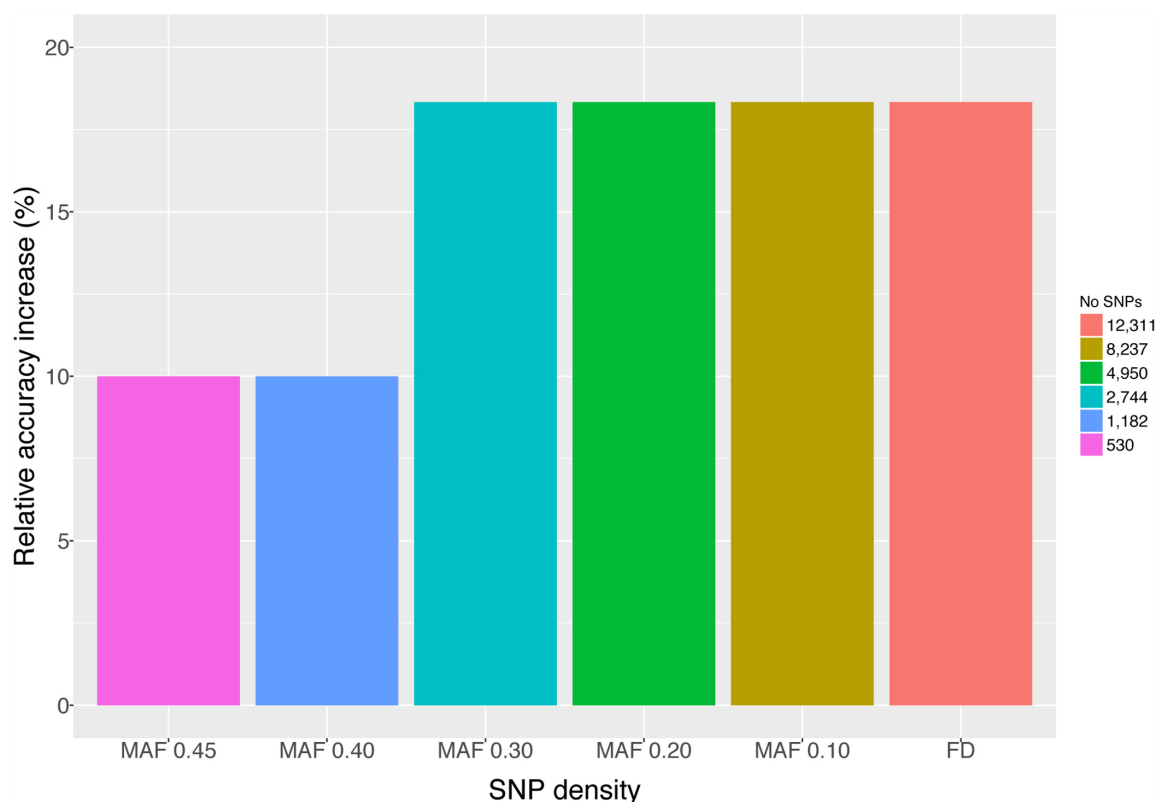
correlation between body weight and SL was 0.94 (SE 0.02), and as such only length data were used for downstream analyses.

### Genome-Wide Association Study (GWAS) – Genomic Selection (GS)

Genome-wide association study and GS were performed using only the SL data. GWAS did not identify SNPs surpassing the genome-wide significant threshold (**Figure 2**), indicating that juvenile growth is likely to be a polygenic trait. Using the cross-validation approach, breeding value prediction accuracy was estimated to be 0.60 (SE 0.03) for PBLUP, as opposed to 0.71 (SE 0.03) with GBLUP. Accuracies obtained through GBLUP using the various reduced SNP densities ranged from 0.66 to 0.71 (**Figure 3**). GBLUP using only SNPs with a minimum MAF of 0.45 (530 SNPs) had approximately a 10% accuracy improvement compared to PBLUP. SNP densities of a minimum MAF of 0.3 had practically the same prediction accuracies as the full data set (**Figure 3**). Estimated bias of (G)EBVs ranged between 0.78 and 1.02. The scenario using SNPs with minimum MAF of 0.45 was found to produce the most biased GEBVs (**Table 2**).

## DISCUSSION

Traditional selective breeding relies on well-documented pedigree, which is relatively straightforward in livestock, but



**FIGURE 3 |** Mean accuracies of GBLUP and PBLUP. GBLUP accuracies were obtained using either all available SNPs (FD; 12,311), SNPs with MAF > 0.10 (8,237), SNPs with MAF > 0.20 (4,950), SNPs with MAF > 0.30 (2,744), SNPs with MAF > 0.40 (1,182), and SNPs with MAF > 0.45 (530).

**TABLE 2 |** Mean accuracy of pedigree BLUP (PBLUP) and GBLUP using SNP of varying MAF obtained from 10 repeats of fivefold cross-validation.

Reps	FD <sup>1</sup> (bias)	MAF <sup>2</sup> > 0.1 (bias)	MAF <sup>3</sup> > 0.2 (bias)	MAF <sup>4</sup> > 0.3 (bias)	MAF <sup>5</sup> > 0.4 (bias)	MAF <sup>6</sup> > 0.45 (bias)	PBLUP (bias)
1st	0.71 (1.05)	0.71 (1.04)	0.71 (1.05)	0.71 (1.06)	0.68 (1.03)	0.66 (0.93)	0.59 (1.02)
2nd	0.70 (1.03)	0.71 (1.02)	0.70 (1.02)	0.70 (1.03)	0.66 (0.99)	0.65 (0.90)	0.59 (1.01)
3rd	0.73 (1.05)	0.73 (1.04)	0.73 (1.05)	0.73 (1.05)	0.67 (0.99)	0.68 (0.92)	0.61 (1.03)
4th	0.69 (0.98)	0.69 (0.97)	0.69 (0.99)	0.69 (0.99)	0.64 (0.95)	0.63 (0.84)	0.60 (1.02)
5th	0.69 (1.04)	0.68 (1.00)	0.70 (0.99)	0.69 (1.02)	0.66 (0.98)	0.66 (0.91)	0.58 (1.02)
6th	0.71 (1.03)	0.71 (1.02)	0.71 (1.03)	0.71 (1.03)	0.67 (1.00)	0.66 (0.90)	0.61 (1.05)
7th	0.70 (1.03)	0.70 (1.03)	0.70 (1.03)	0.70 (1.02)	0.65 (0.97)	0.65 (0.87)	0.60 (1.02)
8th	0.71 (1.04)	0.71 (1.04)	0.71 (1.04)	0.71 (1.05)	0.66 (1.00)	0.66 (0.92)	0.60 (1.03)
9th	0.71 (1.01)	0.71 (1.02)	0.71 (1.03)	0.71 (1.04)	0.67 (1.00)	0.67 (0.92)	0.61 (1.04)
10th	0.70 (1.04)	0.70 (1.02)	0.70 (1.03)	0.70 (1.03)	0.65 (0.98)	0.65 (0.88)	0.60 (1.02)
Mean	0.71 (1.03)	0.71 (1.02)	0.71 (0.98)	0.71 (0.87)	0.66 (0.99)	0.66 (0.90)	0.60 (1.03)

<sup>1</sup> 12,311 SNPs, <sup>2</sup> 8,237 SNPs, <sup>3</sup> 4,950 SNPs, <sup>4</sup> 2,744 SNPs, <sup>5</sup> 1,182 SNPs, <sup>6</sup> 530 SNPs.

more challenging in aquaculture species. Genetic markers have the potential of addressing this issue, facilitating the practical implementation of breeding programs via effective parentage assignment and circumventing the requirement of rearing the fish in separate tanks until tagging is possible (Vandeputte and Haffray, 2014). In the current study, the utility of RAD-seq data for enabling selective breeding for a polygenic trait in common carp was investigated. Using the RAD SNP data, approximately 86% of genotyped animals could be uniquely assigned to parental

pairs. Following pedigree reconstruction, moderate heritability estimates of 0.26 and 0.33 were obtained for juvenile weight and length, respectively. These estimates are in line with the previous heritability estimates of weight/length obtained from juvenile carp (Vandeputte et al., 2004; Ninh et al., 2013; Hu et al., 2017). However, one limitation of the current study lies in the fact that trait measurements were taken in juveniles, and the correlation between growth in early life with growth to harvest size is unknown for this population. Contradicting

evidence is available regarding this, with studies recording high positive phenotypic correlations between growth-related traits at juvenile and harvest stage (Ninh et al., 2013), moderate positive correlations (Vandeputte et al., 2008; Nielsen et al., 2010), or correlations near to zero (Hu et al., 2017). However, common carp juvenile weight and length in the current study is used as an exemplar polygenic trait, with broader implications for the use of genomic data to improve other economically important traits, in particular for those typically not measurable directly on selection candidates (disease resistance or fillet traits, e.g.).

Genetic markers can be a valuable addition to selective breeding, but the optimal strategy for their application depends on the underlying genetic architecture of the trait of interest. Where a trait is primarily controlled by one or several major QTL, it may be most effective to use marker-assisted selection with low-density markers flanking QTL regions. This is the case for resistance to the Infectious Pancreatic Necrosis virus in Atlantic salmon for example, where almost all genetic variation is explained by a single QTL (Houston et al., 2008; Moen et al., 2009). However, most traits of economic importance have a polygenic architecture and GS is likely to be the most effective use of genetic markers to improve these traits. In the current study, the GWAS results implied that the juvenile growth traits were polygenic in nature. Previous studies using linkage analysis have reported QTL related to growth in common carp. A study on a single full-sibling family of common carp detected 14 QTL distributed in five different LGs, including regions associated with the hypothalamic–pituitary–gonadal and GH/IGF-I axis that regulates development, cell-proliferation, energy metabolism, and growth (Peng et al., 2016). Additionally, a study on eight full-sib families of common carp reported 38 growth-related QTL, although no QTL was detected in all of the families (Lv et al., 2016). Therefore, due to the lack of consistent major effect QTL, it is unlikely that MAS will be an effective approach for selecting the best breeding candidates for this and similar economically important traits, and GS is a promising alternative.

The results from the testing of GS in the carp population used in the current study were encouraging, with prediction accuracy obtained through cross-validation analyses using GBLUP being 0.71. This signifies an approximate 18% improvement of accuracy compared to pedigree BLUP, suggesting major potential benefits for selection accuracy and genetic gain for complex economic traits in carp. Our results are in agreement with other studies in aquaculture species where a major benefit to using the genomic models was demonstrated for disease resistance in, e.g., Atlantic salmon (Tsai et al., 2015, 2016), rainbow trout (Vallejo et al., 2017), and gilthead sea bream (Palaikostas et al., 2016). GS benefits from increased sample size of the reference population (Vallejo et al., 2017) indicating that further improvements of prediction accuracies could be expected by increasing the number of genotyped animals in the current study. In typical aquaculture breeding designs, including mass spawning species, where the reference and validation sets are closely related use of genetic markers can be highly effective for capturing within-family genetic variation

(Lillehammer et al., 2013). However, given that prediction accuracy is likely to be highly reliant on genetic relationships, this implies that genomic prediction in distant relatives to the reference population is likely to be substantially more challenging, as observed in terrestrial livestock (Daetwyler et al., 2013).

It is noteworthy that GBLUP resulted in increased prediction accuracies compared to pedigree prediction using relatively sparse (~500 SNPs) genotype data, especially considering the large genome size (~1.8 Gb) of common carp. Similar results were recorded for Atlantic salmon where similar prediction accuracies were obtained from 5K SNPs as for 112K SNPs (Tsai et al., 2015). Both common carp and Atlantic salmon have large genomes, and the effectiveness of genomic prediction at low marker density is again likely to reflect the aforementioned close relationships between the training and validation animals. Nonetheless, this may have economic benefits, since low cost sparse genotyping could be sufficient for improving prediction accuracies in a breeding program. This could be important for driving implementation of GS, since breeding candidates of aquaculture species are of lower economic value compared with livestock, making the application of costly high-density genotyping approaches difficult to justify. Genotype imputation approaches have major potential to drive this genotype density and cost down further, and have already shown significant promise in Atlantic salmon (Tsai et al., 2017). While verification of the results of the current study using harvest size carp (or other economically important complex traits) would be a logical next step, the results of the current study suggest that GS has potential for substantial improvement in prediction accuracy in carp breeding, and RAD-seq is one method of generating the marker data to enable this improvement.

## CONCLUSION

The results from the current study demonstrated that the use of SNP markers generated via RAD-seq is an efficient approach for investigating and potentially improving a polygenic trait in a common carp breeding population. These SNPs enabled pedigree assignment, genetic parameter estimation, GWAS, and GS within a single experiment. GS resulted in improved prediction accuracy versus pedigree approaches even when only relatively sparse marker information was utilized.

## AUTHOR CONTRIBUTIONS

MK, MP, and RH conceived the study and contributed to designing the experimental structure. MK and MP shared on establishing and on-growing the experimental stock, PIT tagging, phenotype data recording, and fin clipping the fish. CP carried out the DNA extractions, RAD library preparation, and sequence data processing. CP and RH carried out parentage assignment and the quantitative genetic analyses. All authors contributed to drafting the manuscript.



## FUNDING

The authors were supported by funding from the European Union's Seventh Framework Programme (FP7 2007-2013) under Grant Agreement No. 613611 (FISHBOOST). MK and MP were also supported by the Ministry of Education Youth, and Sports of the Czech Republic (projects CENAKVA – CZ.1.05/2.1.00/01.0024 and CENAKVA II – LO1205 under the NPU I program) and the Grant Agency of the University of South Bohemia (Project No. 125/2016/Z).

## REFERENCES

- Aguilar, I., Misztal, I., Legarra, A., and Tsuruta, S. (2011). Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128, 422–428. doi: 10.1111/j.1439-0388.2010.00912.x
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi: 10.1371/journal.pone.0003376
- Balon, E. K. (1995). Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers. *Aquaculture* 129, 3–48. doi: 10.1016/0044-8486(94)00227-F
- Barria, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., et al. (2017). Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3*. doi: 10.1534/g3.118.200053 [Epub ahead of print].
- Baxter, S. W., Davey, J. W., Johnston, J. S., Shelton, A. M., Heckel, D. G., Jiggins, C. D., et al. (2011). Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One* 6:e19315. doi: 10.1371/journal.pone.0019315
- Campbell, N. R., LaPatra, S. E., Overturf, K., Towner, R., and Narum, S. R. (2014). Association mapping of disease resistance traits in rainbow trout using restriction site associated DNA sequencing. *G3* 4, 2473–2481. doi: 10.1534/g3.114.014621
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping Loci de novo from short-read sequences. *G3* 1, 171–182. doi: 10.1534/g3.111.000240
- Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G., and Hickey, J. M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193, 347–365. doi: 10.1534/genetics.112.147983
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. doi: 10.1534/genetics.112.143313
- Dong, Z., Nguyen, N. H., and Zhu, W. (2015). Genetic evaluation of a selective breeding program for common carp *Cyprinus carpio* conducted from 2004 to 2014. *BMC Genet.* 16:94. doi: 10.1186/s12863-015-0256-2
- Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., and Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods Mol. Biol.* 772, 157–178. doi: 10.1007/978-1-61779-228-1-9
- FAO (2015). *FishStat Database*. Available at: <http://faostat.fao.org/site/629/default.aspx>
- Ferdosi, M. H., Kinghorn, B. P., van der Werf, J. H. J., Lee, S. H., and Gondro, C. (2014). hspase: an R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. *BMC Bioinformatics* 15:172. doi: 10.1186/1471-2105-15-172

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00082/full#supplementary-material>

**TABLE S1** | Parentage and phenotypic information of juvenile common carp.

**TABLE S2** | Detailed information of identified SNP through RADseq.

**TABLE S3** | Linkage map of common carp.

- Gjedrem, T. (2000). Genetic improvement of cold-water fish species. *Aquac. Res.* 31, 25–33. doi: 10.1046/j.1365-2109.2000.00389.x
- Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447. doi: 10.2307/2529430
- Hervé, P., and Dandine-Roulland, C. (2018). *Gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models Version 1.5 from CRAN*. Available at: <https://rdrr.io/cran/gaston/>
- Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W., and Implementing, Genomic Selection in CGIAR Breeding Programs Workshop Participants. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* 49, 1297–1303. doi: 10.1038/ng.3920
- Houston, R. D., Davey, J. W., Bishop, S. C., Lowe, N. R., Mota-Velasco, J. C., Hamilton, A., et al. (2012). Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* 13:244. doi: 10.1186/1471-2164-13-244
- Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al. (2008). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in atlantic salmon (*Salmo salar*). *Genetics* 178, 1109–1115. doi: 10.1534/genetics.107.082974
- Hu, X., Li, C., Shang, M., Ge, Y., Jia, Z., Wang, S., et al. (2017). Inheritance of growth traits in Songpu mirror carp (*Cyprinus carpio* L.) cultured in Northeast China. *Aquaculture* 477, 1–5. doi: 10.1016/j.aquaculture.2017.04.031
- Hulata, G. (1995). A review of genetic improvement of the common carp (*Cyprinus carpio* L.) and other cyprinids by crossbreeding, hybridization and selection. *Aquaculture* 129, 143–155. doi: 10.1016/0044-8486(94)00244-I
- Janssen, K., Chavanne, H., Berentsen, P., and Komen, H. (2017). Impact of selective breeding on European aquaculture. *Aquaculture* 472, 8–16. doi: 10.1016/j.aquaculture.2016.03.012
- Kocour, M., Mauger, S., Rodina, M., Gela, D., Linhart, O., and Vandeputte, M. (2007). Heritability estimates for processing and quality traits in common carp (*Cyprinus carpio* L.) using a molecular pedigree. *Aquaculture* 270, 43–50. doi: 10.1016/j.aquaculture.2007.03.001
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lillehammer, M., Meuwissen, T. H. E., and Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45:39. doi: 10.1186/1297-9686-45-39
- Lv, W., Zheng, X., Kuang, Y., Cao, D., Yan, Y., and Sun, X. (2016). QTL variations for growth-related traits in eight distinct families of common carp (*Cyprinus carpio*). *BMC Genet.* 17:65. doi: 10.1186/s12863-016-0370-9
- Meuwissen, T., Hayes, B., and Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1, 221–237. doi: 10.1146/annurev-animal-031412-103705
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. H. (2002). “BLUPF90 and related programs (BGF90),” in *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, 21–22.

- Moav, R., and Wohlfarth, G. (1976). Two-way selection for growth rate in the common carp (*Cyprinus carpio* L.). *Genetics* 82, 83–101.
- Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368. doi: 10.1186/1471-2164-10-368
- Nielsen, H. M., Ødegård, J., Olesen, I., Gjerde, B., Ardo, L., Jeney, G., et al. (2010). Genetic analysis of common carp (*Cyprinus carpio*) strains. I: genetic parameters and heterosis for growth traits and survival. *Aquaculture* 304, 14–21. doi: 10.1016/j.aquaculture.2010.03.016
- Ninh, N. H., Ponzoni, R. W., Nguyen, N. H., Woolliams, J. A., Taggart, J. B., McAndrew, B. J., et al. (2013). A comparison of communal and separate rearing of families in selective breeding of common carp (*Cyprinus carpio*): responses to selection. *Aquaculture* 40, 152–159. doi: 10.1016/j.aquaculture.2013.06.005
- Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., and Meuwissen, T. H. E. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5:402. doi: 10.3389/fgene.2014.00402
- Palaïokostas, C., Ferrareso, S., Franch, R., Houston, R. D., and Bargelloni, L. (2016). Genomic prediction of resistance to Pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3* 58, 3693–3700. doi: 10.1534/g3.116.035220
- Palti, Y., Vallejo, R. L., Gao, G., Liu, S., Hernandez, A. G., Rexroad, C. E., et al. (2015). Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated DNA sequencing. *PLoS One* 10:e0138435. doi: 10.1371/journal.pone.0138435
- Peng, W., Xu, J., Zhang, Y., Feng, J., Dong, C., Jiang, L., et al. (2016). An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (*Cyprinus carpio*). *Sci. Rep.* 6:26693. doi: 10.1038/srep26693
- Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., and Auvinen, P. (2013). Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* 29, 3128–3134. doi: 10.1093/bioinformatics/btt563
- Robledo, D., Palaïokostas, C., Bargelloni, L., Martínez, P., and Houston, R. (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* doi: 10.1111/raq.12193 [Epub ahead of print].
- Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. doi: 10.1186/1471-2164-15-478
- Sonesson, A. K., and Meuwissen, T. H. E. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Sel. Evol.* 41:37. doi: 10.1186/1297-9686-41-37
- Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., et al. (2016). Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* 48:47. doi: 10.1186/s12711-016-0226-9
- Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., et al. (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* 16:969. doi: 10.1186/s12864-015-2117-9
- Tsai, H.-Y., Matika, O., Edwards, S. M., Antolin-Sánchez, R., Hamilton, A., Guy, D. R., et al. (2017). Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3*, 1377–1383. doi: 10.1534/g3.117.040717
- Vallejo, R. L., Leeds, T. D., Fragomeni, B. O., Gao, G., Hernandez, A. G., Misztal, I., et al. (2016). Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: insights on genotyping methods and genomic prediction models. *Front. Genet.* 7:96. doi: 10.3389/fgene.2016.00096
- Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* 49:17. doi: 10.1186/s12711-017-0293-6
- Vandeputte, M. (2003). Selective breeding of quantitative traits in the common carp (*Cyprinus carpio*): a review. *Aquat. Living Resour.* 16, 399–407. doi: 10.1016/S0990-7440(03)00056-1
- Vandeputte, M., and Haffray, P. (2014). Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* 5:432. doi: 10.3389/fgene.2014.00432
- Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Rodina, M., et al. (2004). Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). *Aquaculture* 235, 223–236. doi: 10.1016/j.aquaculture.2003.12.019
- Vandeputte, M., Kocour, M., Mauger, S., Rodina, M., Launay, A., Gela, D., et al. (2008). Genetic variation for growth at one and two summers of age in the common carp (*Cyprinus carpio* L.): heritability estimates and response to selection. *Aquaculture* 277, 7–13. doi: 10.1016/j.aquaculture.2008.02.009
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., et al. (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat. Genet.* 46, 1212–1219. doi: 10.1038/ng.3098

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Palaïokostas, Kocour, Prchal and Houston. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.